Appendix 8.2

Model Performance Evaluation (TSD Chapter 3)

3.0 MODEL PERFORMANCE EVALUATION

In this Chapter we summarize the CMAQ model performance for the final 2002 36 km Base F base case simulation. This model performance focuses on the ability of the model to predict PM species within the CENRAP region. Details on the model performance are provided in Appendix C. Previously we have documented model performance of interim versions of model base case simulations in reports (Morris et al., 2005) and presentations to the CENRAP Work Groups and POG (e.g., Morris et al., 2006a,b).

3.1 Evaluation Methodology

EPA's integrated ozone, $PM_{2.5}$ and regional haze modeling guidance calls for a comprehensive, multi-layered approach to model performance testing, consisting of the four major components: operational, diagnostic, mechanistic (or scientific) and probabilistic (EPA, 2007). The CMAQ model performance evaluation effort focused on the first two components, namely:

- **Operational Evaluation**: Tests the ability of the model to estimate PM concentrations (both fine and coarse) and the components at PM_{10} and $PM_{2.5}$ including the quantities used to characterize visibility (i.e., sulfate, nitrate, ammonium, organic carbon, elemental carbon, other $PM_{2.5}$, and coarse matter ($PM_{2.5-10}$). This evaluation examines whether the measurements are properly represented by the model predictions but does not necessarily ensure that the model is getting "the right answer for the right reason"; and

- **Diagnostic Evaluation:** Tests the ability of the model to predict visibility and extinction, PM chemical composition including PM precursors (e.g., SOx, NOx, and NH_3) and associated oxidants (e.g., ozone and nitric acid); PM size distribution; temporal variation; spatial variation; mass fluxes; and components of light extinction (i.e., scattering and absorption).

In this final model performance evaluation for the 2002 Typical Base F CMAQ simulation, the operational evaluation has been given the greatest attention since this is the primarily thrust of EPA's modeling guidance. However, we have also examined certain diagnostic features dealing with the model's ability to simulate sub-regional and monthly/diurnal gas phase and aerosol concentration distributions. In the course of the CENRAP and other modeling process numerous diagnostic sensitivity tests were performed to investigate and improve model performance. Key diagnostic tests performed are discussed and the results for the rest are available on the CENRAP modeling website: http://pah.cert.ucr.edu/aqm/cenrap/index.shtml.

3.2 Ambient Air Quality Data used in the Evaluation

The ground-level model evaluation database for 2002 was compiled by the modeling team using several routine and research-grade databases. The first is the routine gas-phase concentration measurements for ozone, SO2, NO_2 and CO archived in EPA's Aerometric Information Retrieval System (AIRS) Air Quality System (AQS) database. Other sources of observed information come from the various PM monitoring networks in the U.S. These include the Interagency Monitoring of Protected Visual Environments (IMPROVE); Clean Air Status and Trends Network (CASTNET); EPA Speciation Trends Network (STN) of $PM_{2.5}$ species; and National Acid Deposition Network (NADP). During the course of the CENRAP modeling, the numerous base case simulations were evaluated across the continental U.S. (e.g., Morris et al., 2005). In this section and in Appendix C we focus our evaluation on model performance within the CENRAP region.

3.2 Operational Model Evaluation Approach

The CENRAP modeling databases will be used to develop the visibility State Implementation Plan (SIP) as required by the Regional Haze Rule (RHR). Accordingly, the primary focus of the operational evaluation in this report is on the six components of fine particulate ($PM_{2.5}$) and coarse mass ($PM_{2.5\text{-}10}$) within the CENRAP region that are used to characterize visibility at Class I areas:

- Sulfate (SO4);
- Particulate Nitrate (NO3);
- Elemental Carbon (EC);
- Organic Mass Carbon (OMC);
- Other inorganic fine particulate (IP or Soil); and
- Coarse Mass (CM).

The model performance for ozone and precursor and product species (e.g., SO_2 and HNO_3) is also evaluated to build confidence that the modeling system is sufficiently reliable to project future-year visibility.

3.3 Model Performance Goals and Criteria

The issue of model performance goals for PM species is an area of ongoing research and debate. For ozone modeling, EPA has established performance goals for 1-hour ozone normalized mean bias and gross error of #±15% and #35%, respectively (EPA, 1991). EPA's draft fine particulate modeling guidance notes that performance goals for ozone should be viewed as upper bounds of model performance that PM models may not be able to always achieve and we should demand better model performance for PM components that make up a larger fraction of the PM mass than those that are minor contributors (EPA, 2001). EPA's final modeling guidance does not list any specific model performance goals for PM and visibility modeling and instead provides a summary of PM model performance across several historical applications that can be used for comparisons if desired. Measuring PM species is not as precise as ozone monitoring. In fact, the differences in measurement techniques for some species likely exceed the more stringent performance goals, such as those for ozone. For example, recent comparisons of the PM species measurements using the IMPROVE and STN measurement technologies found differences of approximately ∀20% (SO4) to ∀50% (EC) (Solomon et al., 2004).

For the CENRAP modeling we have adopted three levels of model performance goals and criteria for bias and gross error as listed in Table 3-1. Note that we are not suggesting that these performance goals be adopted as guidance. Rather, we are just using them to frame and put the PM model performance into context and to facilitate model performance intercomparison across episodes, species, models and sensitivity tests.

Table 3-1. Model performance goals and criteria used to assist in interpreting modeling results.

Fractional Bias	Fractional Error	Comment
#∀15%	#35%	Ozone model performance goal for which PM model performance would be considered good – note that for many PM species measurement uncertainties may exceed this goal.
#∀30%	#50%	Proposed PM model performance goal that we would hope each PM species could meet
#∀60%	#75%	Proposed PM criteria above which indicates potential fundamental problems with the modeling system.

As noted in EPA's PM modeling guidance, less abundant PM species should have less stringent performance goals (EPA, 2001; 2007). Accordingly, we are also using performance goals that are a continuous function of average concentrations, as proposed by Dr. James Boylan at the Georgia Department of Natural Resources (GA DNR), that have the following features (Boylan, 2004):

- Asymptotically approaching proposed performance goals or criteria (i.e., the ∀30%/50% and ∀60%/75% bias/error levels listed in Table 3-1) when the mean of the observed concentrations are greater than 2.5 ug/m^3.

- Approaching 200% error and ∀200% bias when the mean of the observed concentrations are extremely small.

Bias and error are plotted as a function of average concentrations. As the mean concentration approach zero, the bias performance goal and criteria flare out to ∀200% creating a horn shape, hence the name "Bugle Plots". Dr. Boylan has defined three Zones of model performance: Zone 1 meets the ∀30%/50% bias/error performance goal and is considered "good" model performance; Zone 2 lies between the ∀30%/50% performance goal and ∀60%/75% performance criteria and is an area where concern for model performance is raised; and Zone 3 lies above the ∀60%/75% performance criteria and is an area of questionable model performance.

3.4 Key Measures of Model Performance

Although we have generated numerous statistical performance measures (see Table C-2 in Appendix C) that are available on the CENRAP modeling website, when comparing model performance across months, subdomains, networks, grid resolution, models, studies, etc. it is useful to have a few key measurement statistics to be used to facilitate the comparisons. It is also useful to have a subset of the 2002 year that can represent the entire year so that a more focused evaluation can be conducted. We have found that the Mean Fractional Bias and Mean Fractional Gross Error appear to be the most consistent descriptive measure of model performance (Morris et al., 2004b; 2005). The Fractional Bias and Error normalize by the average of the observed and predicted value (see Table C-2) because it provides descriptive power across different magnitudes of the model and observed concentrations and is bounded by -200% to +200%. This is in contrast to the normalized bias and error (as recommended for ozone performance goals, EPA, 1991) that is normalized by just the observed value so can "blow up" to infinity as the observed value approaches zero. In Appendix C we perform a focused evaluation of model performance for PM and gaseous species and four months of the 2002 year that are used to represent the seasonal variation in performance:

- January
- April
- July
- October

Scatter plots of model predictions and observations for each PM species is presented for each of the four months along with performance statistics and predicted and observed time series plots at each CENRAP Class I area. Summary plots of monthly fractional bias and error are also presented.

3.5 Operational Model Performance Evaluation

A summary of the operational evaluation is presented below. Just the monthly fractional bias performance metrics for each PM species using bar charts and Bugle Plots are presented in this section. The reader is referred to Appendix C for the complete model performance evaluation.

3.5.1 Sulfate (SO4) Model Performance

Figure 3-1 compares the monthly SO4 fractional bias and error across the CENRAP region for the IMPROVE, STN and CASTNet monitoring networks. An under-prediction bias is clearly evident the first 8-10 months of the year. This underestimation bias is greatest across the CASTNet network which persists throughout the year and is least for the STN network where it disappears by August-September. For the IMPROVE network, the SO4 fractional bias is < ±20% for the first 2 and last 3 months of the year and ranges from -30% to -50% for the late Spring and Summer months.

Figure 3-1 also includes a Bugle Plot of monthly SO4 fractional bias and error statistics and compares them against the proposed PM model performance goal and criteria (see Table 3-1).

For the STN network, it appears that SO4 performance for all months achieves the proposed PM model performance goal. For the IMPROVE network, approximately half of the months achieve the proposed PM performance goal with the other half exceeding the goal but within the performance criteria. Across the CASTNet network, most months exceed the proposed goal and are within the criteria. Although the CASTNet fractional bias for some months is right at the performance criteria (≤±60%). With the exception of two IMPROVE months, all of the monthly SO4 fractional error performance statistics achieve the proposed PM model performance goal.

3.5.2 Nitrate (NO3) Model Performance

Monthly NO3 model performance across he CENRAP region is characterized by a summer underestimation and winter overestimation bias (Figure 3-2). The summer underestimation bias is more severe exceeding -100%, whereas the winter overestimation bias is approximately 50%. So based on statistics alone, it appears the summer underestimation bias is a bigger concern than the winter overestimation bias. However, the Bugle Plots in the bottom part of Figure 3-2 shows that the summer underestimation bias occurs when NO3 is very low and is not an important component of PM and visibility impairment. These summer values occur in the flared horn part of the Bugle Plot and in fact the summer NO3 performance mostly achieves the model performance goal and always achieves the performance criteria. Whereas the winter overstated NO3 performance mostly doesn't meet the performance goal and there are even some months/networks that don't meet the performance criteria.

3.5.3 Organic Matter Carbon (OMC) Model Performance

The OMC monthly fractional bias across IMPROVE and STN sites in the CENRAP region are shown in Figure 3-3. The bias performance for OMC at the IMPROVE sites is quite good throughout the year with values generally within ±20%, albeit with a slight winter overestimation and summer underestimation bias. At the urban STN sites, the model exhibits an underestimation bias throughout the year that ranges from -20% to -50%. The urban underestimation of OMC is a fairly common occurrence and suggests there may be missing sources of organic aerosol emissions.

The good performance of the model for OMC at the IMPROVE sites is also reflected in the Bugle Plot (Figure 3-3, bottom) with the bias and error achieving the proposed PM model performance goal for all months of the year. At the STN sites, however, the OMC bias falls between the proposed PM model performance goal and criteria, with error right at the goal for most months.

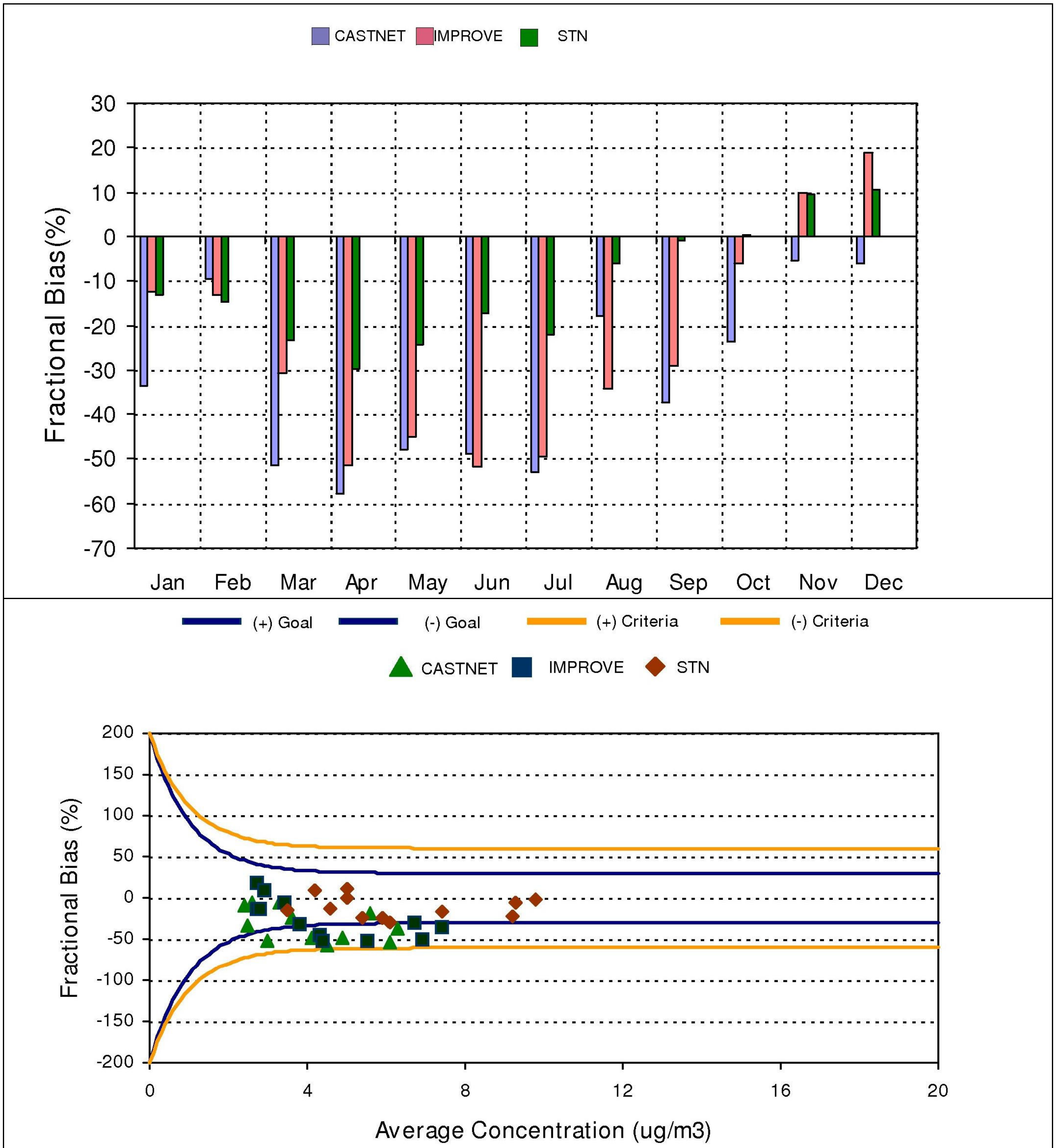
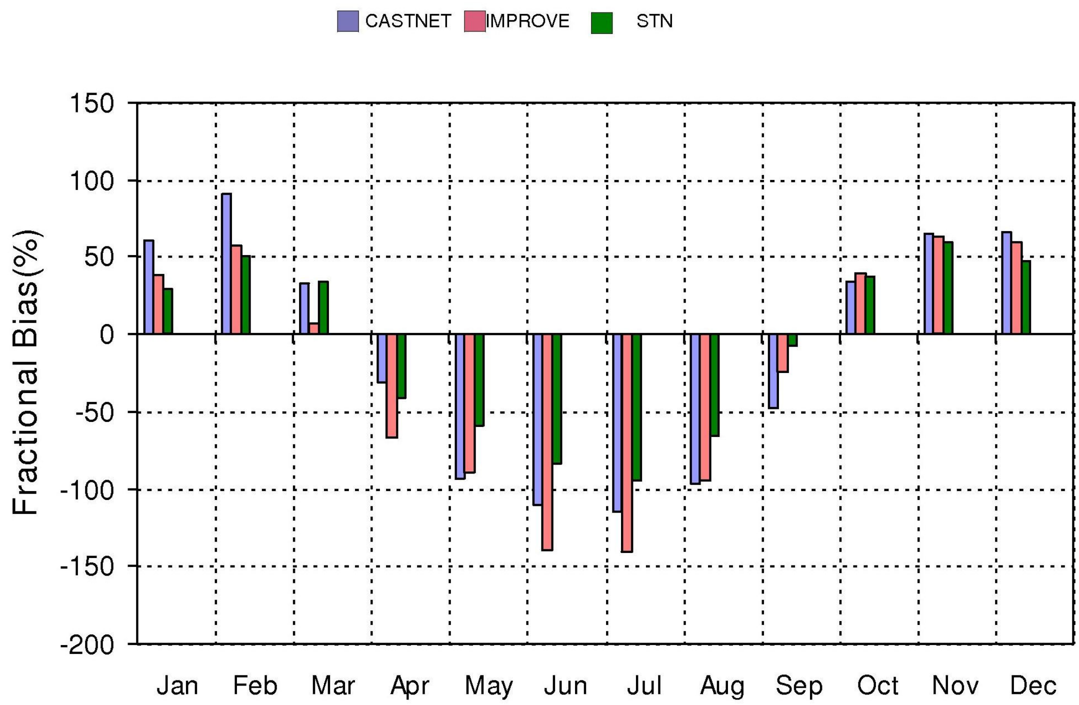


Figure 3-1. Monthly fractional bias (%) for sulfate (SO4) across the CENRAP region for the CMAQ 2002 36 km Base F base case simulation.



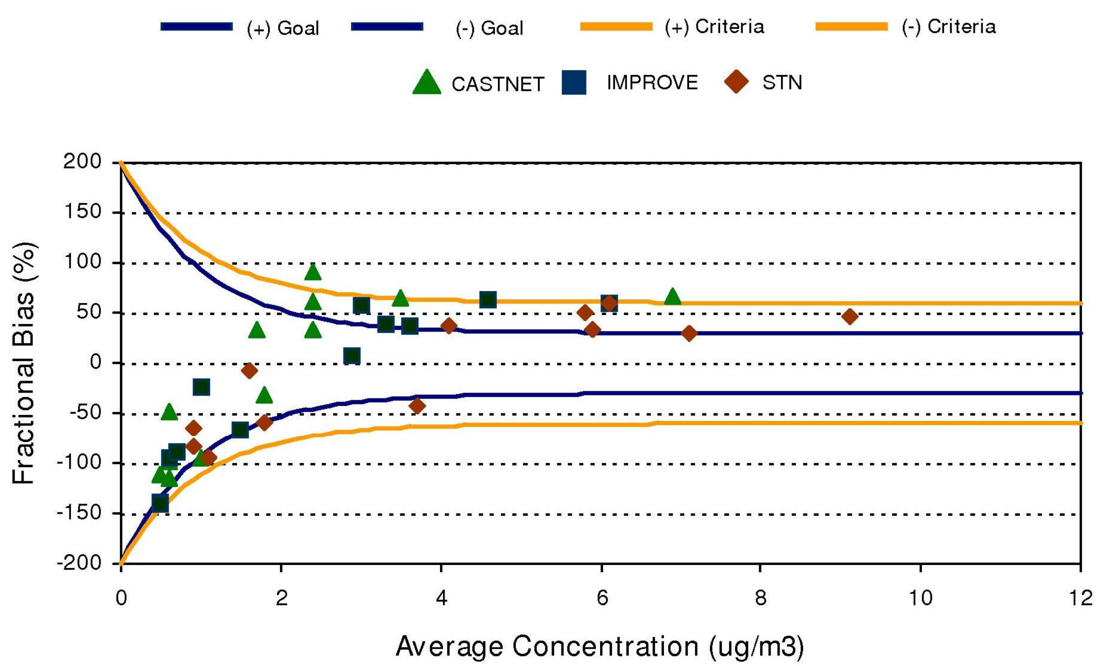


Figure 3-2. Monthly fractional bias (%) for nitrate (NO3) across the CENRAP region for the CMAQ 2002 36 km Base F base case simulation.

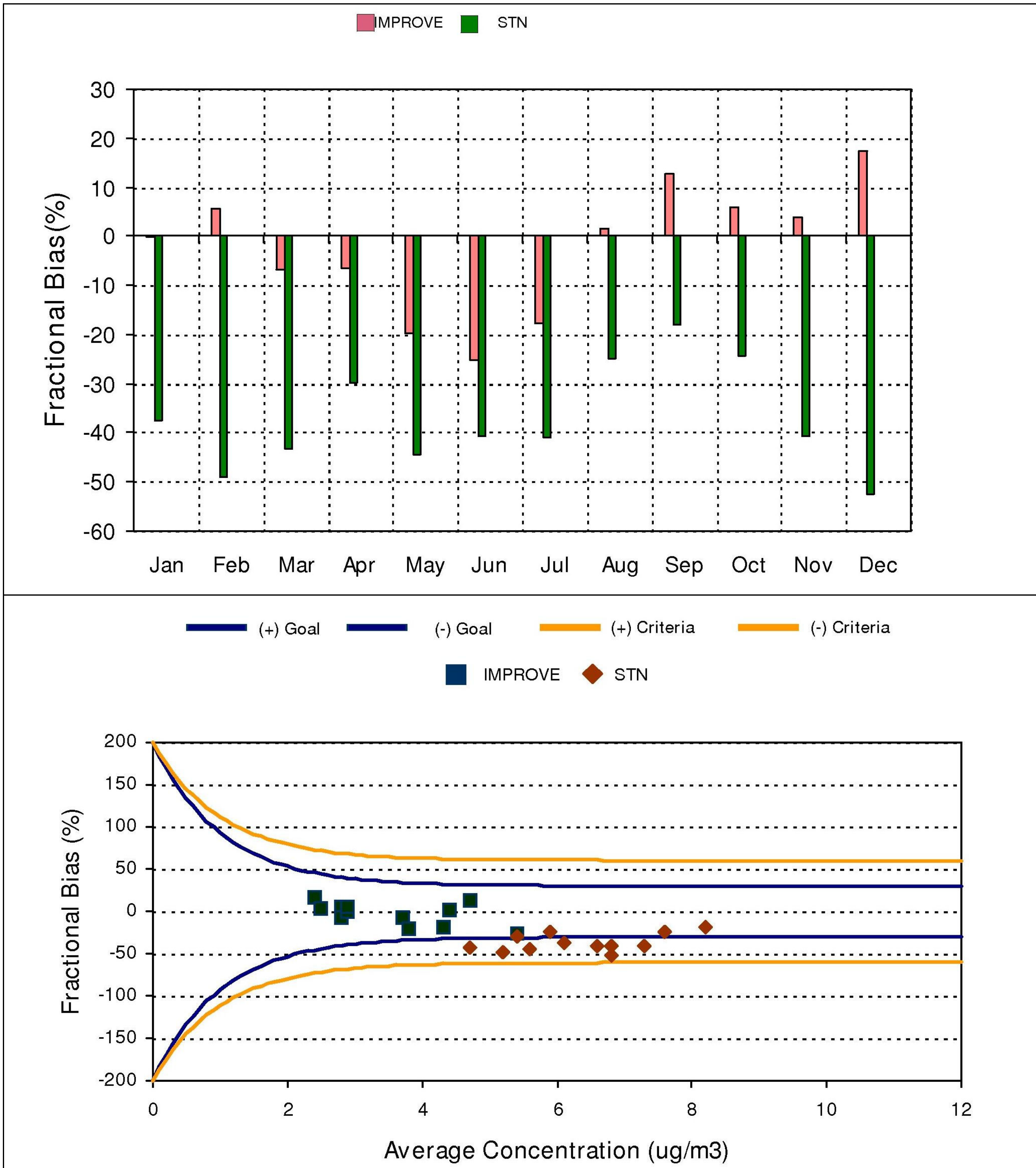


Figure 3-3. Monthly fractional bias (%) for organic matter carbon (OMC) across the CENRAP region for the CMAQ 2002 36 km Base F base case simulation.

3.5.4 Elemental Carbon (EC) Model Performance

The monthly average bias and error for EC across the IMPROVE and STN monitors in the CENRAP region are shown in Figure 3-4. The STN network exhibits low bias year round, whereas the IMPROVE monitoring network exhibits a large under-prediction bias in the summer months (-40% to -60%) and much lower EC bias in the winter. The Bugle Plot puts the EC performance in context. The low EC concentrations at the IMPROVE sites results in bias values in the horn of the Bugle Plot. Thus, EC bias and error performance achieves the proposed PM performance goal for all months of the year.

3.5.5 Other $PM_{2.5}$ (Soil) Model Performance

Figure 3-5 displays the monthly variation in the Soil fractional bias using IMPROVE measurements in the CENRAP region. During the winter months, the model exhibits a very large (> 100%) overestimation bias. With the exception of July, the summer monthly bias is toward a slight over-prediction but generally less than 20% with errors of 60% to 80%. The July underestimation bias appears to be driven by impacts of high Soil values from wind blown dust events (e.g., see July 2002 discussion in Appendix C). The Bugle Plot indicates that the summer Soil performance achieves the PM performance goal, a few months in the Spring/Fall period fall between the performance goal and criteria and the winter Soil performance exceeds the model performance criteria. Thus, the Soil performance is a cause for concern.

3.5.6 Coarse Mass (CM) Model Performance

The monthly average fractional bias values for CM are shown in Figure 3-6. In the winter the under-prediction bias is typically in the -60% to -80% range. In the late Spring and Summer the under-prediction bias ranges from -120% to -160%. As this under-prediction bias is nearly systematic, then the errors are the same magnitude as the bias.

The Bugle Plots clearly show that the CM model performance is a problem. The monthly bias exceeds both the performance goal and criteria for almost every month of the year. The error criteria are also exceeded for all months of the year.

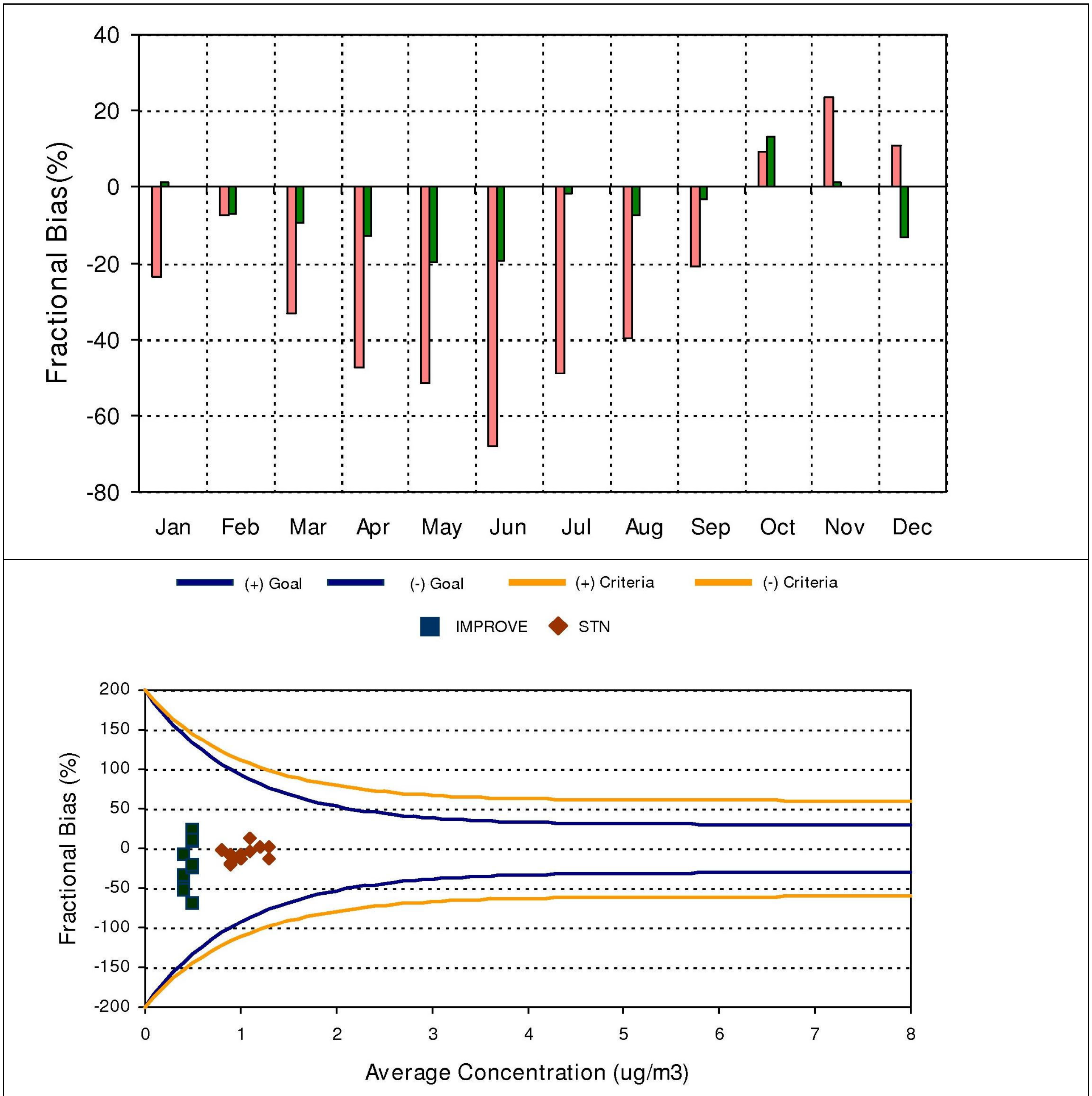
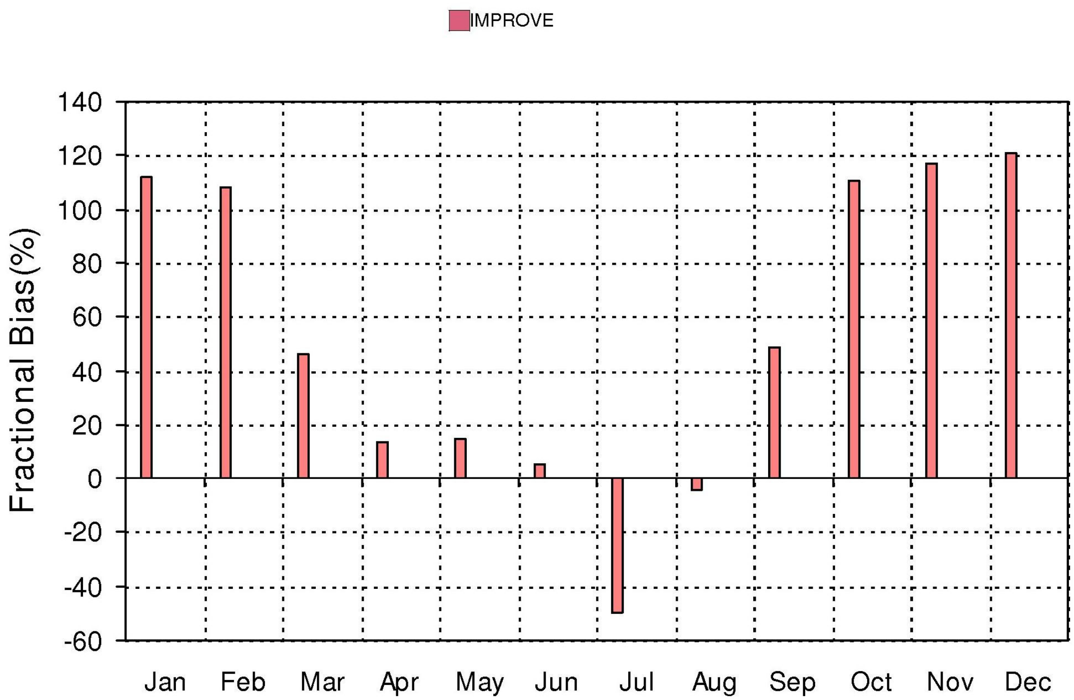


Figure 3-4. Monthly fractional bias (%) for elemental carbon (EC) across the CENRAP region for the CMAQ 2002 36 km Base F base case simulation.



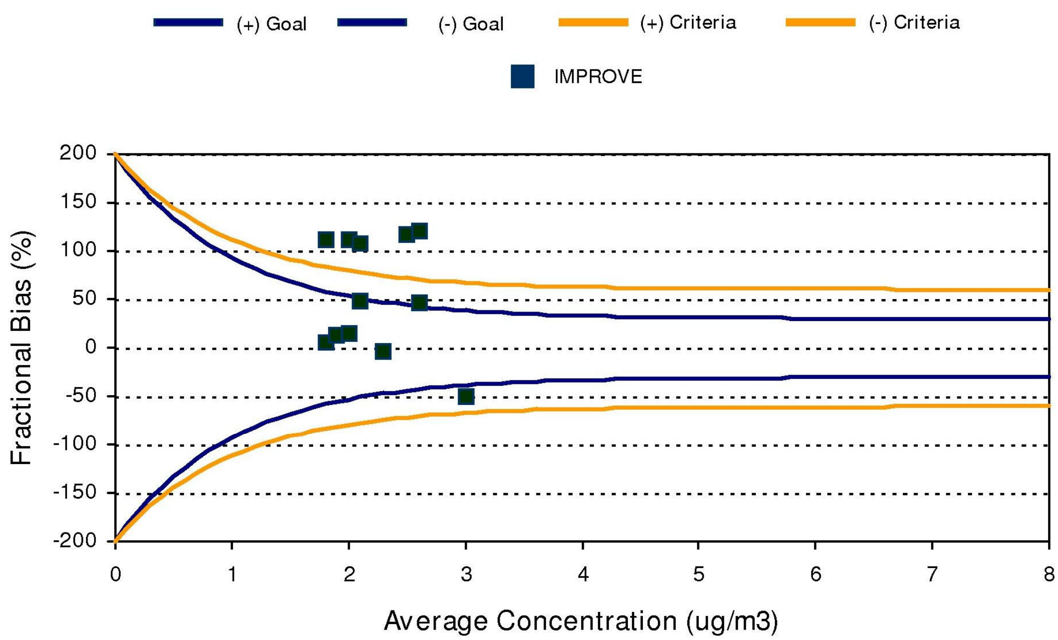


Figure 3-5. Monthly fractional bias (%) for other $PM_{2.5}$ (Soil) across the CENRAP region for the CMAQ 2002 36 km Base F base case simulation.

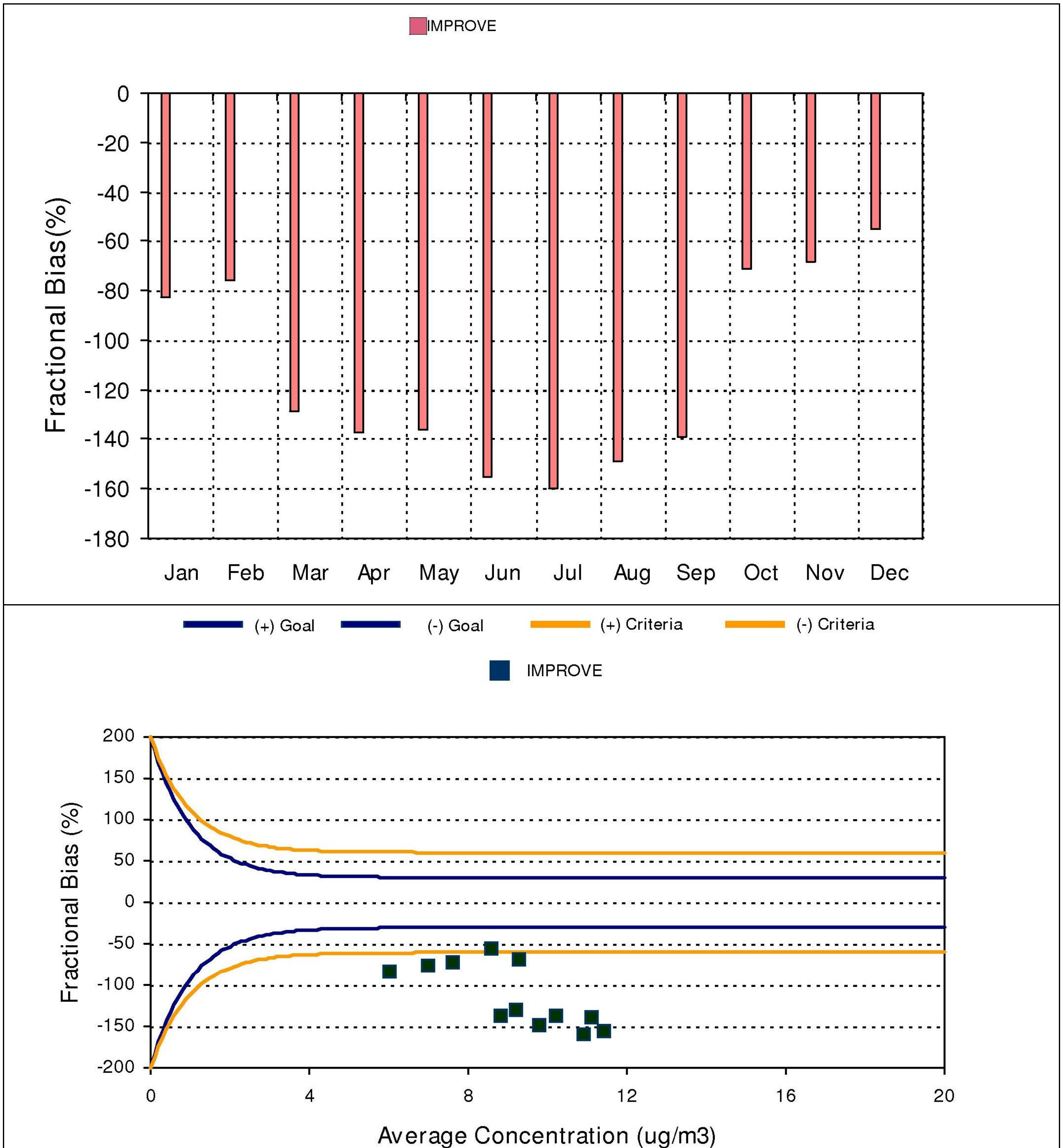


Figure 3-6. Monthly fractional bias (%) for coarse mass (CM) across the CENRAP region for the CMAQ 2002 36 km Base F base case simulation.

3.6 Diagnostic Model Performance Evaluation

The CASTNet and AQS networks also measure gas-phase species that are PM precursor or related species. The diagnostic evaluation of the 2002 36 km Base F CMAQ base case simulation for these compounds and the four seasonal months are presented in Appendix C. The displays for January are provided below as an example; the reader is referred to Appendix C for the rest of the monthly displays.

The CASTNet network measures weekly average samples of SO2, SO4, NO2, HNO3, NO3 and NH4. The AQS network collects hourly measurements of SO2, NO2, O3 and CO. A comparison of the SO2 and SO4 performance provides insight into whether the SO4 formation rate may be too slow or fast. For example, if SO4 is underestimated and SO2 is overestimated that may indicate too slow chemical conversion rates. Analyzing the performance for SO4, HNO3, NO3, Total NO3 and NH4 provides insight into the equilibrium of these species. For example, if Total NO3 performs well but HNO3 and NO3 do not, then there may be issues associated with the partitioning between the gaseous and particle phases of nitrate. Causes for incorrect HNO3/NO3 partitioning could include inadequate ammonia emissions and/or poorly characterized meteorological conditions (e.g., temperature).

3.6.1 Diagnostic Model Performance in January 2002

In January, SO2 is overstated across both the CASTNet and AQS sites with fractional bias values of 38% (Figure 3-7) and 31% (Figure 3-8), respectively. SO4 is understated by -34% across the CASTNet monitors (Figure 3-7) and -12% and -13% for the IMPROVE and STN networks (Figure C-4a). Wet SO4 deposition is also overstated in January (+40%, Figure C-4a). Given that SO2 emissions are well characterized, these results suggest that the January SO4 underestimation may be partly due to understated transformation rates of SO2 to SO4 and overstated wet SO4 deposition.

Total NO3 is overestimated by 35% on average across the CASTNet sites in the CENRAP region in January (Figure 3-7). HNO3 is underestimated (-34%) and particle NO3 is overestimated (+61%) suggesting there are gas/particle equilibrium issues. An analysis of the time series of the four CASTNet stations reveals that NO3, HNO3 and NH4 performance is actually very reasonable at the west Texas site and the HNO3 underestimation and NO3 overestimation bias is coming from the east Kansas, central Arkansas and northern Minnesota CASTNet sites (see Figure C-3 for site locations). One potential contributor for this performance problem could be overstated NH3 emissions. However, the Total NO3 overestimation bias suggests that the model estimated NOx oxidation rate may be too high in January.

The SO2, NO2, O3 and CO performance across the AQS sites in January is shown in Figure 3-8. The AQS monitoring network is primarily an urban-oriented network so it is not surprising that the model is underestimating concentrations of primary emissions like NO2 (-5%) and, particularly, CO (-67%) when a 36 km grid is used. Ozone is also underestimated on average, especially the maximum values above 60 ppb.

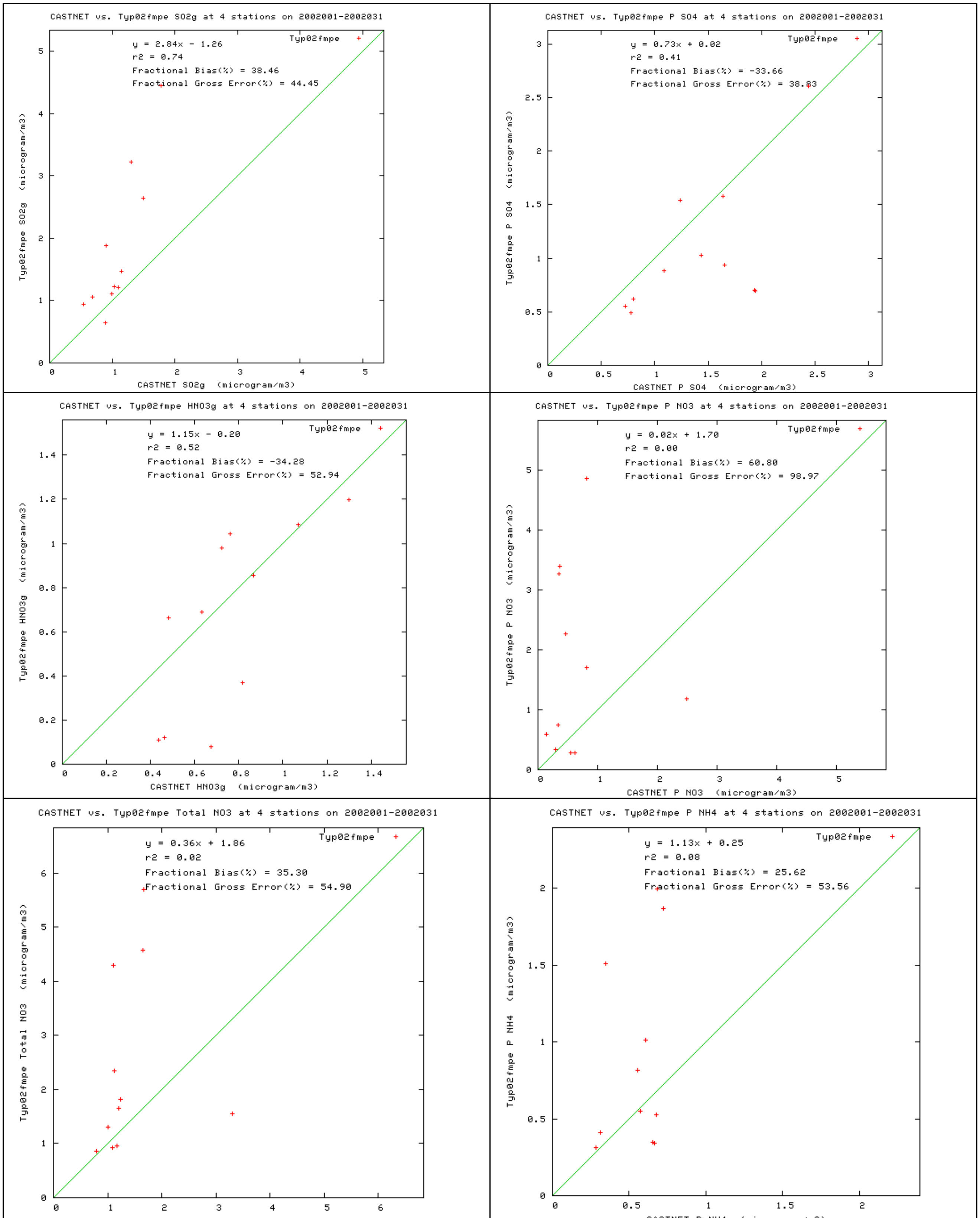


Figure 3-7. January 2002 performance at CENRAP CASTNet sites for SO2 (top left), SO4 (top right), HNO3 (middle left), NO3 (middle right), Total NO3 (bottom left) and NH4 (bottom right).

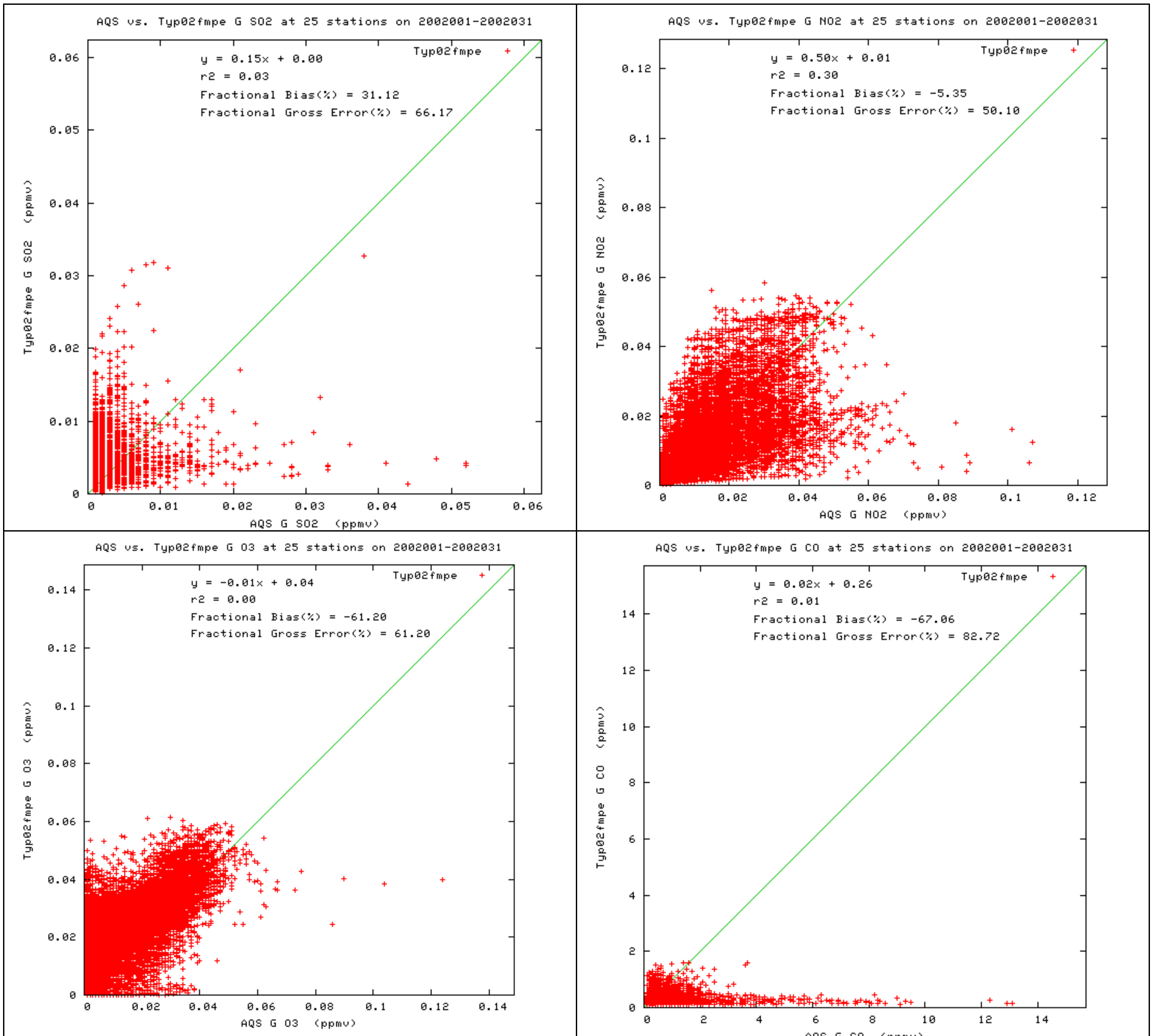


Figure 3-8. January 2002 performance at CENRAP AQS sites for SO2 (top left), NO2 (top right), O3 (bottom left) and CO (bottom right).

3.6.2 Diagnostic Model Performance In April

In April there is an average SO2 overestimation bias across the CASTNet (+15%) and underestimation bias across the AQS (-10%) networks (Figures C-42 and C-43). SO4 is underestimated across all networks by -30% to -58% (Figure C-5a). The wet SO4 deposition bias is near zero. Both SO2 and SO4 are underestimated at the west Texas CASTNet monitor in April suggesting SO2 emissions in Mexico are likely understated.

The HNO3 performance in April is interesting with almost perfect agreement except for 5 modeled-observed comparisons that drives the average under-prediction bias of -29% (Figure C-42). On Julian Day 102 there is high HNO3 at the MN, KS and OK CASTNet sites that is not captured by the model. Given that HNO3, NO3 and Total NO3 are all underestimated by about the same amount (-30%), then part of the underestimation bias is likely due to too slow oxidation of NOx.

There is a lot of scatter in the NO2 and O3 performance that is more or less centered on the 1:1 line of perfect agreement with bias values of -8% and -21%, respectively (Figure C-43). CO is underestimated by -72% with the model unable to predict CO concentrations above 1 ppm due to the use of the coarse 36 km grid spacing. Mobile sources produce a vast majority of the CO emissions so AQS monitors for CO compliance are located near roadways, which are not simulated well using a 36 km grid.

3.6.3 Diagnostic Model Performance In July

In July SO2 is slightly underestimated across the CASTNet (-5%) and AQS (-12%) networks (Figures C-44 and C-45) and SO4 is more significantly underestimated across all networks (-22% to -53%, Figure C-6a). Since wet SO4 is also underestimated it is unclear the reasons for why all sulfur species are underestimated.

The nitrate species are also all underestimated with the Total NO3 bias (-56%) being between the HNO3 bias (-35%) and NO3 bias (-115%). The modeled NO3 values are all near zero with little correlation with the observations, whereas the observed HNO3 and Total NO3 is tracked well with correlation coefficients of 0.74 and 0.76. These results suggest that the July NO3 model performance problem is partly due to insufficient formation of Total NO3 but mainly due to too little incorrect partitioning of the Total NO3 into the particle NO3.

Again there is lots of scatter in the AQS NO2 scatter plot for July (Figure C-45) resulting in a low bias (0%) but high error (65%). Ozone performance also exhibits a low bias (-15%) and error (20%), but the model is incapable of simulating ozone above 100 ppb. Although CO performance in July is better than the previous months, it still has a large underestimation bias (-82%).

3.6.4 Diagnostic Model Performance In October

SO2 is overstated in October across the CASTNet (+28%) and AQS (+33%) sites (Figures C-46 and C-47). Although SO4 is understated across the CASTNet sites (-24%), the bias across the IMPROVE (-6%) and STN (0%) sites are near zero (Figure C-7a).

Performance for HNO3 is fairly good with a low bias (+12%) and error (30%). But NO3 is overstated (+34%) leading to an overstatement of Total NO3 (+37%). The overstatement of NO3 leads to an overstatement of NH4 as well (Figure C-46)

As seen in the other months, NO2 exhibits a lot of scatter resulting in a low correlation (0.22) and high error (61%) but low bias (12%). The model tends to under-predict the high and over-predict the low O3 observations resulting in a -29% bias and low correlation coefficient. CO is also under-predicted (-76%) for the reasons discussed previously.

3.7 Performance at CENRAP Class I Areas for the Worst and Best 20 Percent Days

In this section, and in section C.5 of Appendix C, we present the results of the model performance evaluation at each of the CENRAP Class I areas for the worst and best 20 percent days. Performance on these days is critical since they are the days used in the 2018 visibility projections discussed in Chapter 4. For each Class I area we compared the predicted and observed extinction of the worst and best 20 percent days below. In Appendix C the PM species-specific extinction is also compared for the worst 20 percent days.

3.7.1 Caney Creek (CACR) Arkansas

The ability of the CMAQ model to estimate visibility extinction at the CACR Class I area on the 2002 worst and best 20 percent days is provide in Figures 3-9 and C-48. On most of the worst 20 percent days at CACR total extinction is dominated by SO4 extinction with some extinction due to OMC. On four of the worst 20 percent days extinction is dominated by NO3. The average extinction across the worst 20 percent days is underestimated by -33% (Figure 3-9), which is primarily due to a -51% underestimation of SO4 extinction combined with a 6% overestimation of NO3 extinction (Figure C-48). Performance for OMC extinction at CACR on the worst 20 percent days is pretty good with a -20% bias and 36% error, EC extinction is systematically underestimated, Soil extinction has low bias (-19%) buts lots of scatter and high error (74%), while CM extinction is greatly underestimated (bias of -153%).

On the best 20 percent days at CACR the observed extinction ranges from 20 to 40 Mm^{-1}, whereas then modeled extinction has a much larger range from 15 to 120 Mm^{-1}. Much of the modeled overestimation of total extinction on the best 20% days (+44% bias) is due to NO3 overestimation (+94% bias).

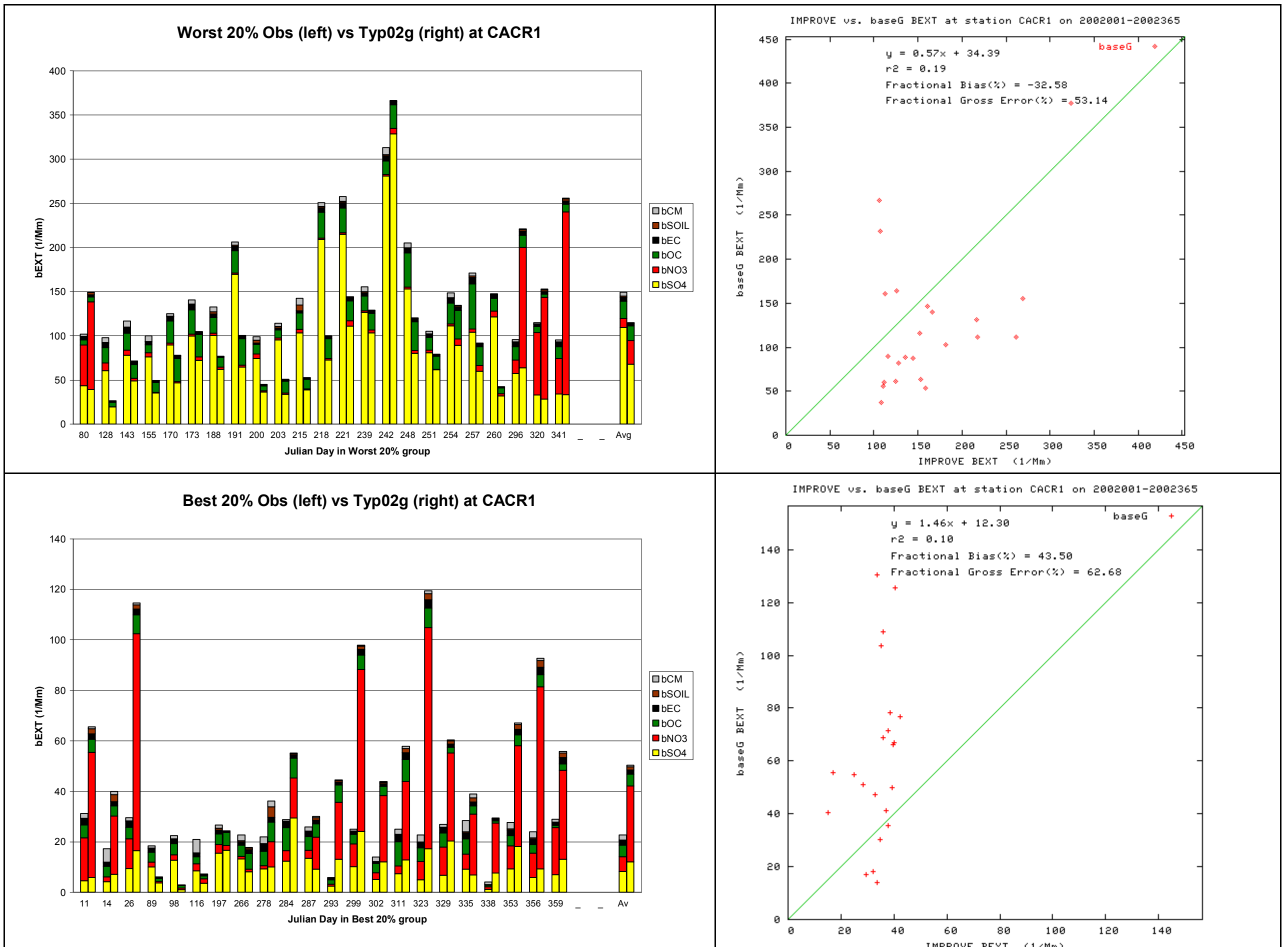


Figure 3-9. Daily extinction model performance at Caney Creek (CACR), Arkansas for the worst (top) and best (bottom) 20 percent days during 2002.

3.7.2 Upper Buffalo (UOBU) Arkansas

Model performance at the UPBU Class I area for the worst and best 20 percent days is shown in Figures 3-10 and C-49. On most of the worst 20 percent days at UPBU visibility impairment is dominated by SO4, although there are also two high NO3 days. The model underestimates the average of the total extinction on the worst 20 percent days at UPBU by -40% (Figure 3-10), which is due to an underestimation of extinction due to SO4, OMC and CM by, respectively, -46%, -33% and -179%.

On the best 20 percent days at UPBU, the model performs reasonably well with a low bias (2%) and error (42%). But again the model has a much wider range in extinction values across the best 20 percent days (15 to 120 Mm^{-1}) than observed (20 to 45 Mm^{-1}). There are five days in which the modeled NO3 over-prediction is quite severe and when those days are removed the range in the modeled and observed extinction on the best 20 percent days is quite similar, although the model gets much cleaner on the very cleanest modeled days.

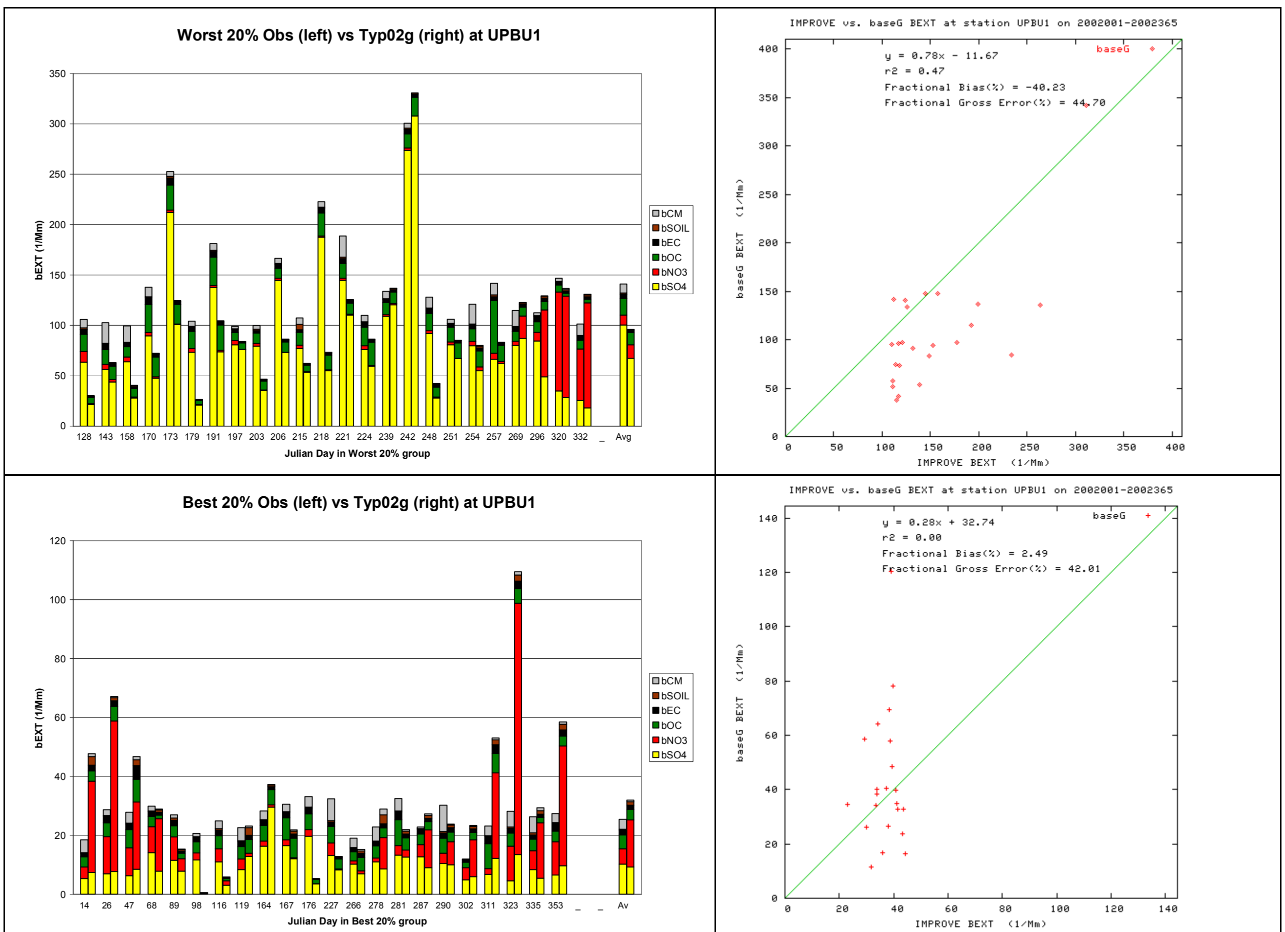


Figure 3-10. Daily extinction model performance at Upper Buffalo (UPBU), Arkansas for the worst (top) and best (bottom) 20 percent days during 2002.

3.7.3 Breton Island (BRET), Louisiana

The observed total extinction on the worst 20 percent days at Breton Island is underestimated by -71% (Figure 3-11), which is due to an underestimation of each component of extinction (Figure C-50) by from -50% to -70% (SO4, OMC and Soil) to over -100% (EC and CM). The observed extinction on the worst 20 percent days ranges from 90 to 170 Mm^{-1}, whereas the modeled values drop down to as low as approximately 15 Mm^{-1}. On the best 20 percent days the range of the observed and modeled extinction is similarly (roughly 10 to 50 Mm^{-1}) that results in a reasonably low bias (-22%), but there is little agreement on which days are higher or lower resulting in a lot of scatter and high error (54%).

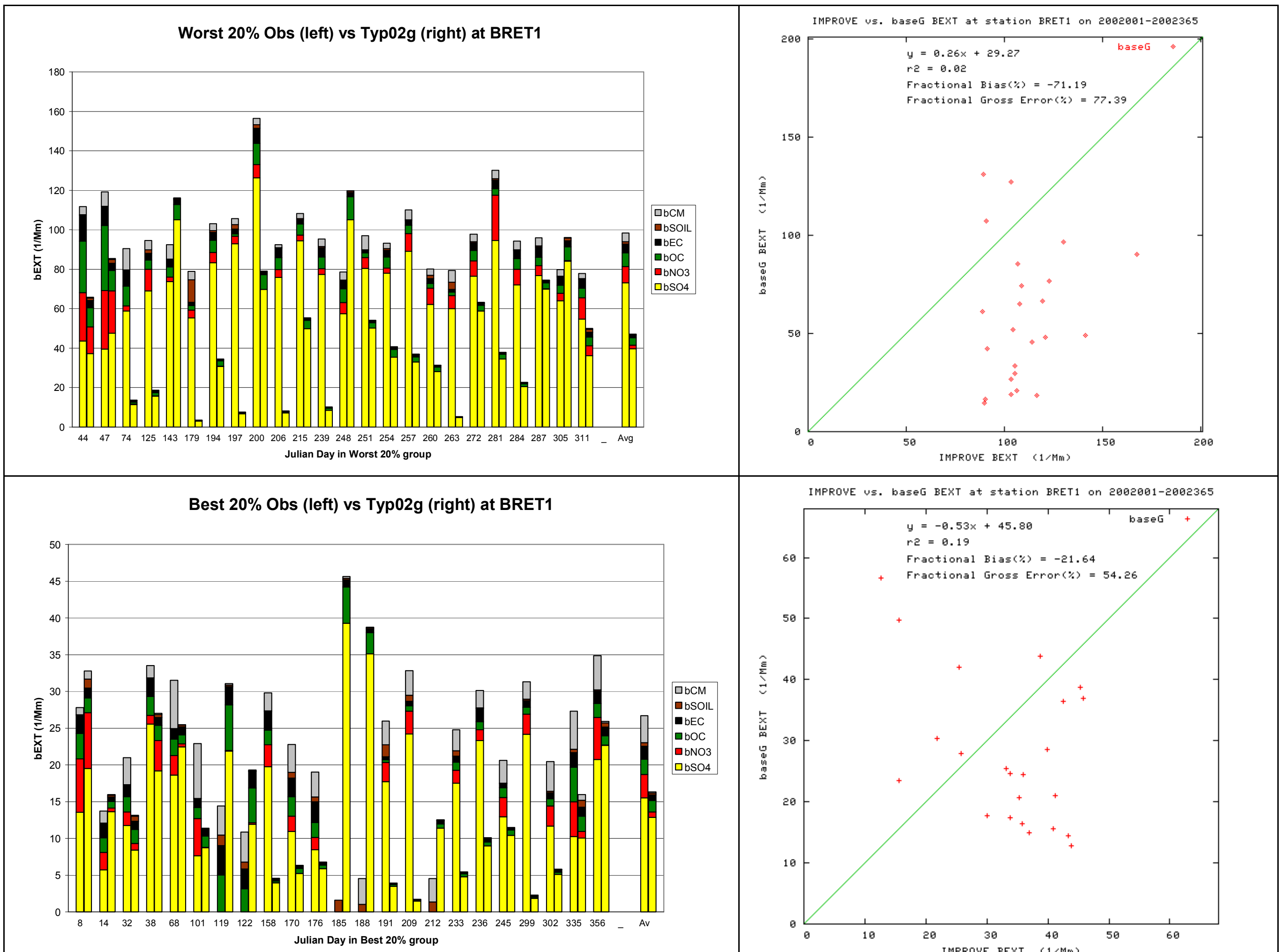


Figure 3-11. Daily extinction model performance at Breton Island (BRET), Louisiana for the worst (top) and best (bottom) 20 percent days during 2002.

3.7.4 Boundary Waters (BOWA), Minnesota

There are three types of days during the worst 20 percent days at BOWA, SO4 days, OMC days and NO3 days (Figure 3-12). The two high OMC days are likely fire impact events that the model captures to some extent on one day and not on the other. On the five high (> 20 Mm^{-1}) NO3 extinction days the model predicts the observed extinction well on three days and overestimates by a factor of 3-4 on the other two high NO3 days. SO4 in underestimate by -43% on average across the worst 20 percent days at BOWA.

With the exception of two days, the model reproduces the total extinction for the best 20 percent days at BOWA quite well with a bias and error value of +14% and 22% (Figure 3-12). Without these two days, the modeled and observed extinction both range between 15 and 25 Mm^{-1}.

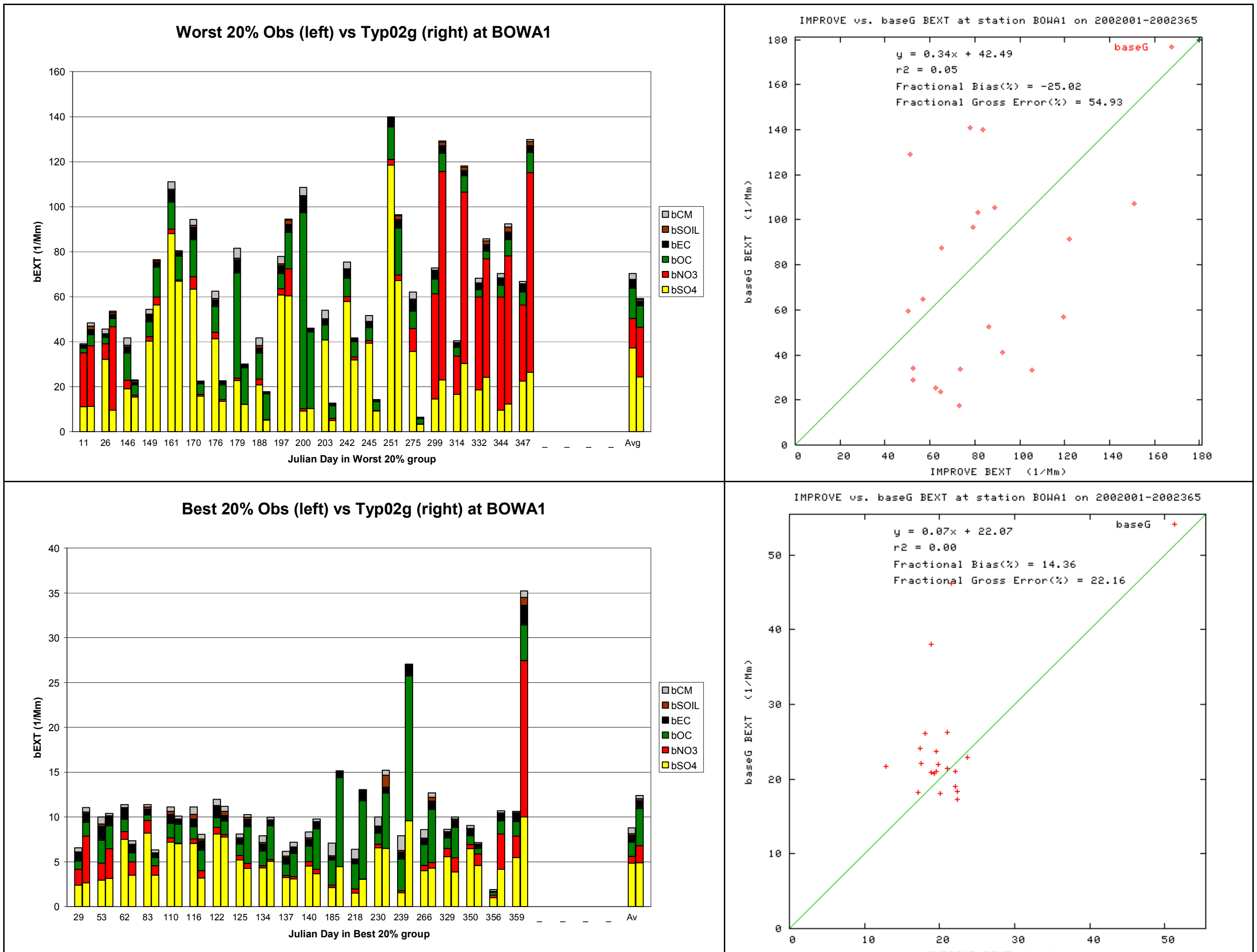


Figure 3-12. Daily extinction model performance at Boundary Waters (BOWA), Minnesota for the worst (top) and best (bottom) 20 percent days during 2002.

3.7.5 Voyageurs (VOYA) Minnesota

VOYA is also characterized by SO4, NO3 and OMC days (Figure 3-13). Julian Days 179 and 200 are high OMC days that were also high OMC days at BOWA again indicating impacts from fires in the area that is not fully captured by the model. SO4 and NO3 extinction is fairly good and, without the fire days, OMC performance looks good as well (Figure C-52). On the best 20 percent days there is one day the modeled extinction is much higher than observed and a few others that are somewhat higher, but for most of the best 20 percent days the modeled extinction is comparable to the observed values.

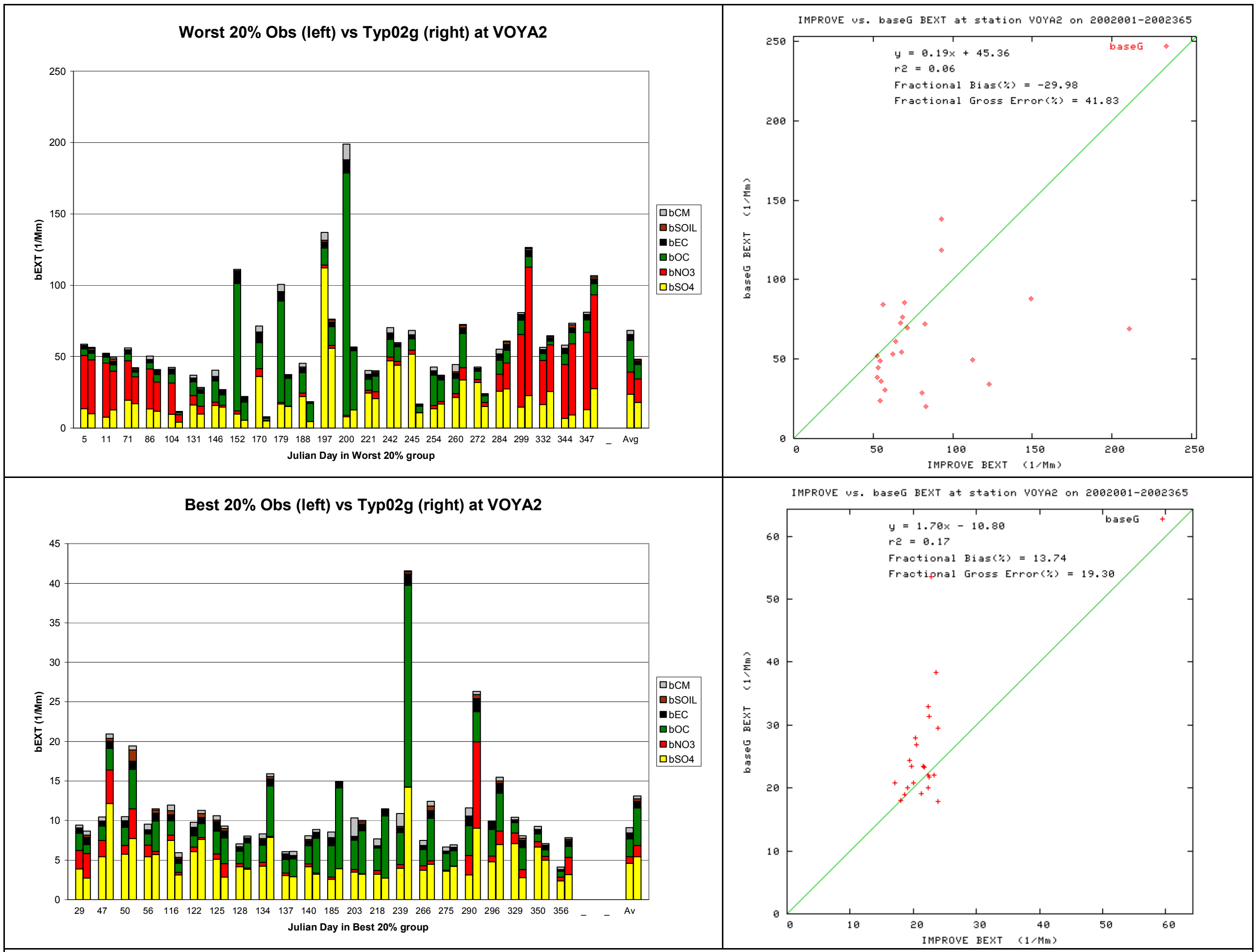


Figure 3-13. Daily extinction model performance at Voyageurs (VOYA), Minnesota for the worst (top) and best (bottom) 20 percent days during 2002.

3.7.6 Hercules Glade (HEGL) Missouri

On most of the worst 20 percent days at HEGL the observed extinction ranges from 120 to 220 Mm^{-1} whereas model extinction ranging from 50 to 170 Mm^{-1} (Figure 3-14). However, there is one extreme day with extinction approaching 400 Mm^{-1} that the model does a very good job in replicating. Over all the days there is a modest underestimation bias in SO4 (-39%) and OMC (-39%) extinction, larger underestimation bias in EC (-62%) and CM (-118%) extinction and overestimation bias in Soil (+30%) extinction (Figure C-53).

On the best 20 percent days there is one day where the model overstates the observed extinction by approximately a factor of four and a handful of other days that the model overstates the extinction by a factor of 2 or so, but most of the days both the model and observed extinction sites are around 40 Mm^{-1} ±10 Mm^{-1}. On the best 20 percent days when the observed extinction is overstated it is due to overstatement of the NO3.

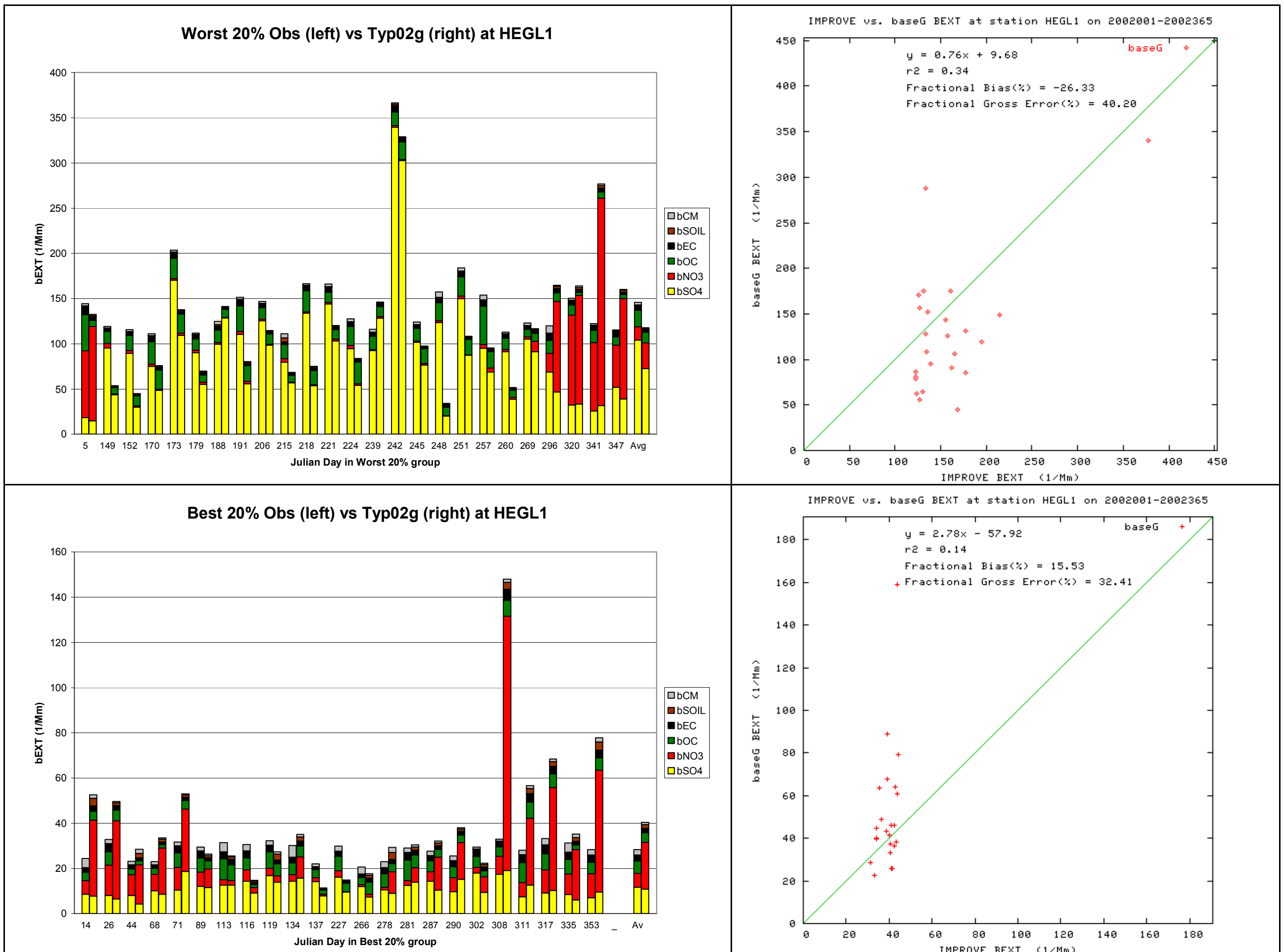


Figure 3-14. Daily extinction model performance at Hercules Glade (HEGL), Missouri for the worst (top) and best (bottom) 20 percent days during 2002.

3.7.7 Mingo (MING) Missouri

The worst 20 percent days at Ming are mainly high SO4 days with a few high NO3 days that the model reproduces reasonably well resulting in low bias (+10%) and error (38%) for total extinction (Figure 3-15). The PM species specific performance is fairly good with low bias for SO4 (+4%), good agreement with NO3 on high NO3 days except for one day, low OMC (+23%) and EC (+3%) bias and larger bias in EC (+37%) and CM (-105%) extinction (Figure C-54).

For the best 20 percent days, there is one day the model is way to high due to overstated NO3 extinction and a few other days the model overstates the observed extinction that is usually due to overrated NO3, but on most of the best 20 percent days the modeled extinction is comparable to the observed values. This results in low bias (+12%) and error (36%) for total extinction at MING for the best 20 percent days.

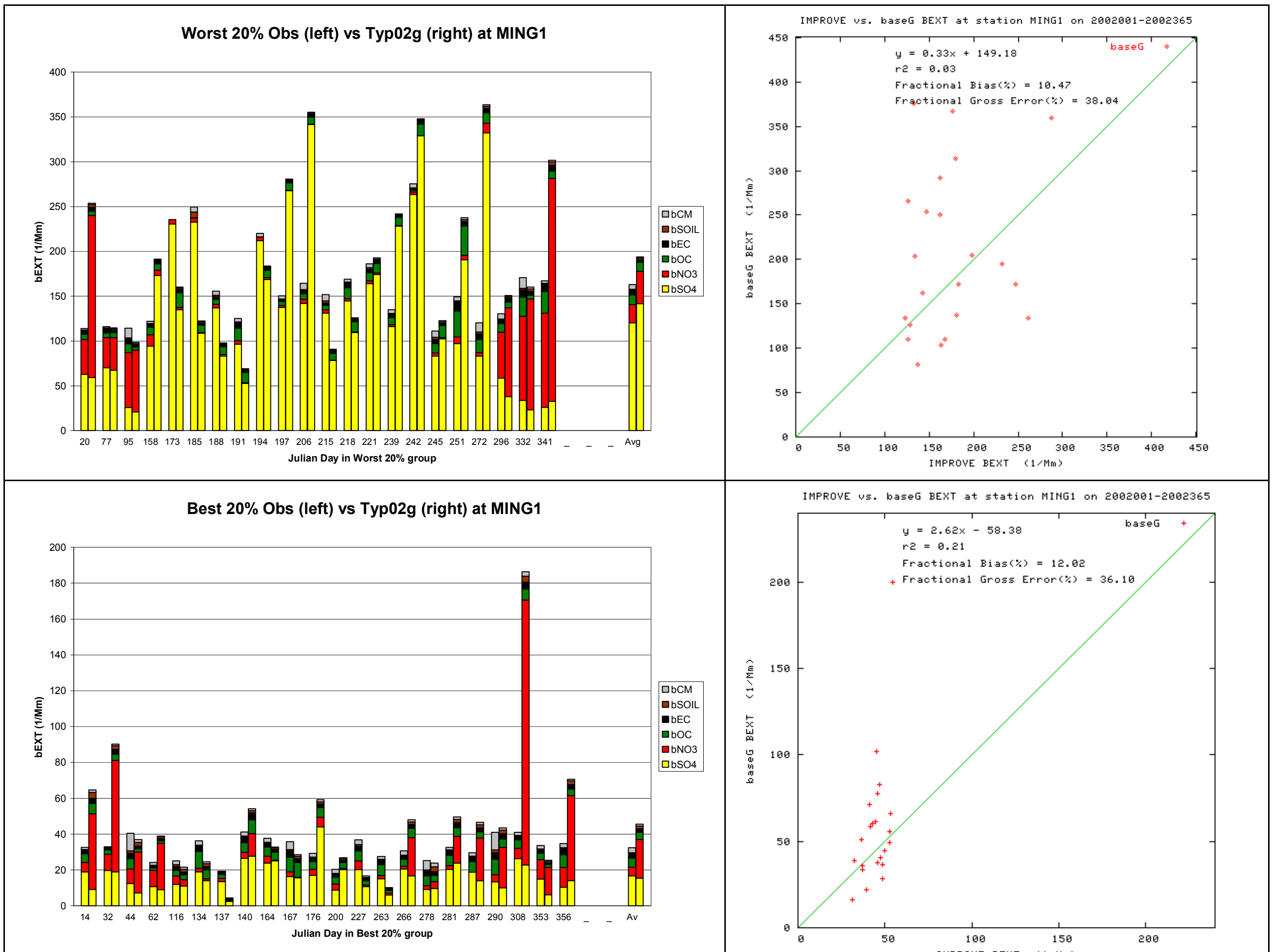


Figure 3-15. Daily extinction model performance at Mingo (MING), Missouri for the worst (top) and best (bottom) 20 percent days during 2002.

3.7.8 Wichita Mountains (WIMO), Oklahoma

With the exception of an over-prediction on day 344 due to NO3, observed total extinction on the worst 20 percent days at WIMO is understated with a bias of -42% (Figure 3-16) that is primarily due to an underestimation of extinction due to SO4 (-48%) and OMC (-69%) (Figure C-55).

CMAQ total extinction performance for the average of the best 20 percent days at WIMO is characterized by an overestimation bias (+21%) on most days that is primarily due to NO3 over-prediction on several days. Again the modeled range of extinction on the best 20 percent days (12-60 Mm^{-1}) is much greater than observed (20-35 Mm^{-1}).

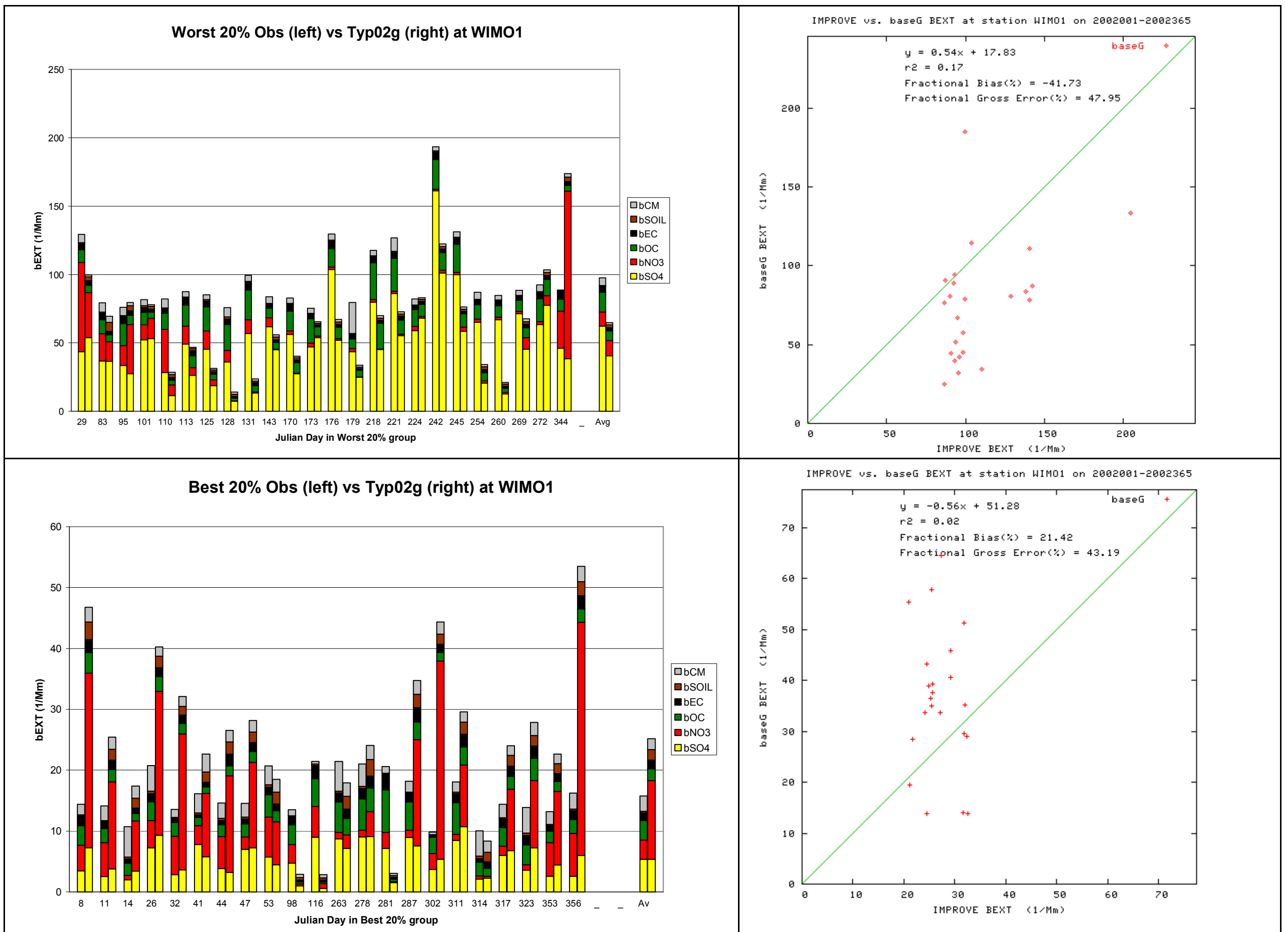


Figure 3-16. Daily extinction model performance at Wichita Mountains (WIMO), Oklahoma for the worst (top) and best (bottom) 20 percent days during 2002.

3.7.9 Big Bend (BIBE) Texas

The observed extinction on the worst 20 percent days at BIBE is under-predicted on almost every day resulting in a fractional bias value of -72% (Figure 3-17). Every component of extinction is underestimated on average for the worst 20 percent days (Figure C-56) with the underestimation bias ranging from -24% (OMC) to -162% (CM). SO4 extinction, that typically represents the largest component of the total extinction is understated by -94%.

The model does a better job in predicting the total extinction at BIBE for the best 20 percent days with average fractional bias and error values of +13% and 19% (Figure 3-17). With the exception of one day that the observed extinction is overestimated by approximately a factor of 2, the modeled and observed extinction on the best 20 percent days at BIBE are both within 12 to 25 Mm^{-1}. However, there are some mismatches with the components of extinction with the model estimating much lower contributions due to Soil and CM.

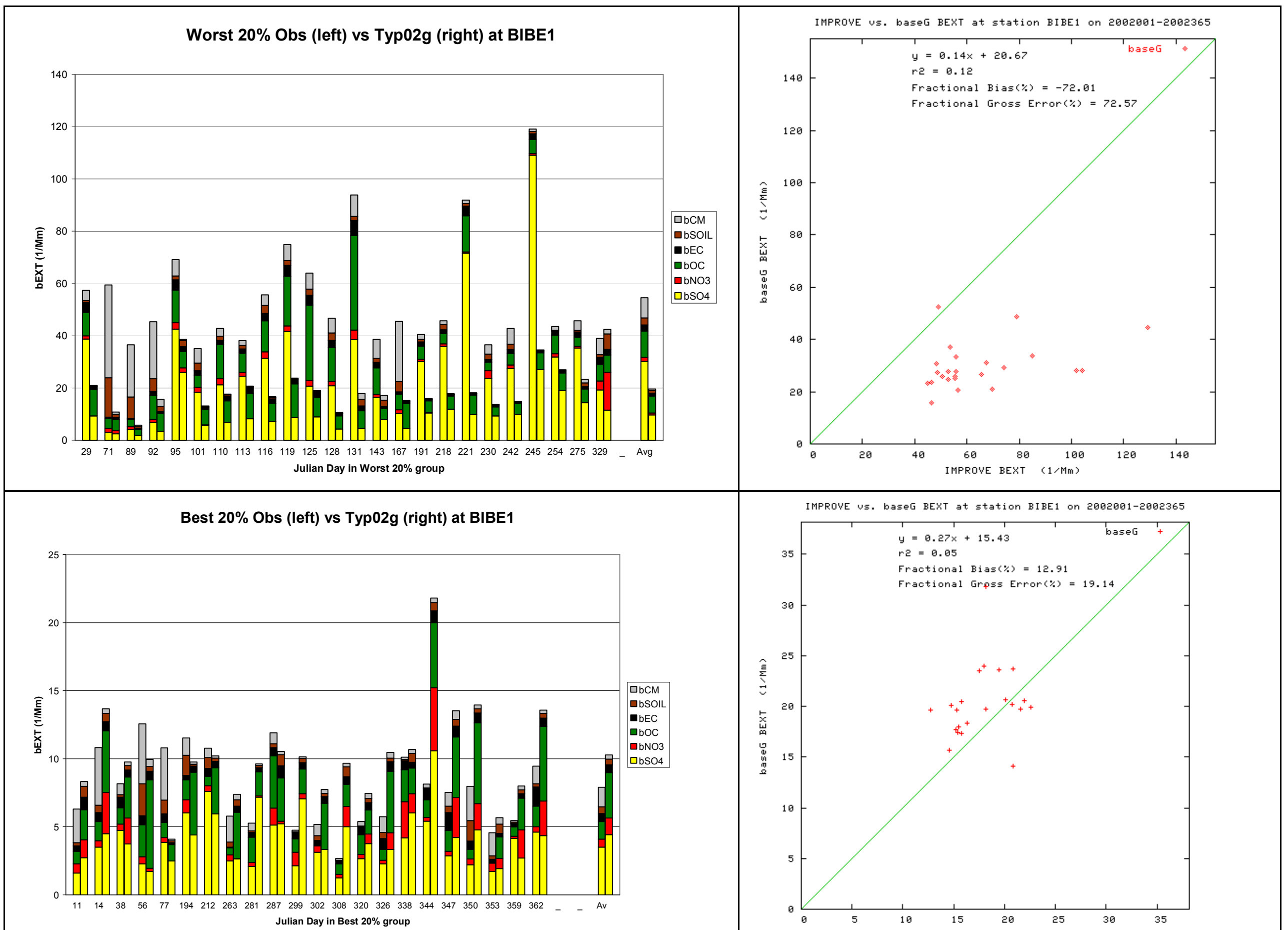


Figure 3-17. Daily extinction model performance at Big Bend (BIBE), Texas for the worst (top) and best (bottom) 20 percent days during 2002.

3.7.10 Guadalupe Mountains (GUMO) Texas

Most of the worst 30 percent days at GUMO are dust days with high Soil and CM that is not at all captured by the model (Figure 3-18). Extinction due to Soil and CM on the worst 20 percent days is underestimated by -105% and -191%, respectively (Figure C-57). Better performance is seen on the best 20 percent days with bias and error for total extinction of 8% and 21%, but the model still understates Soil and CM.

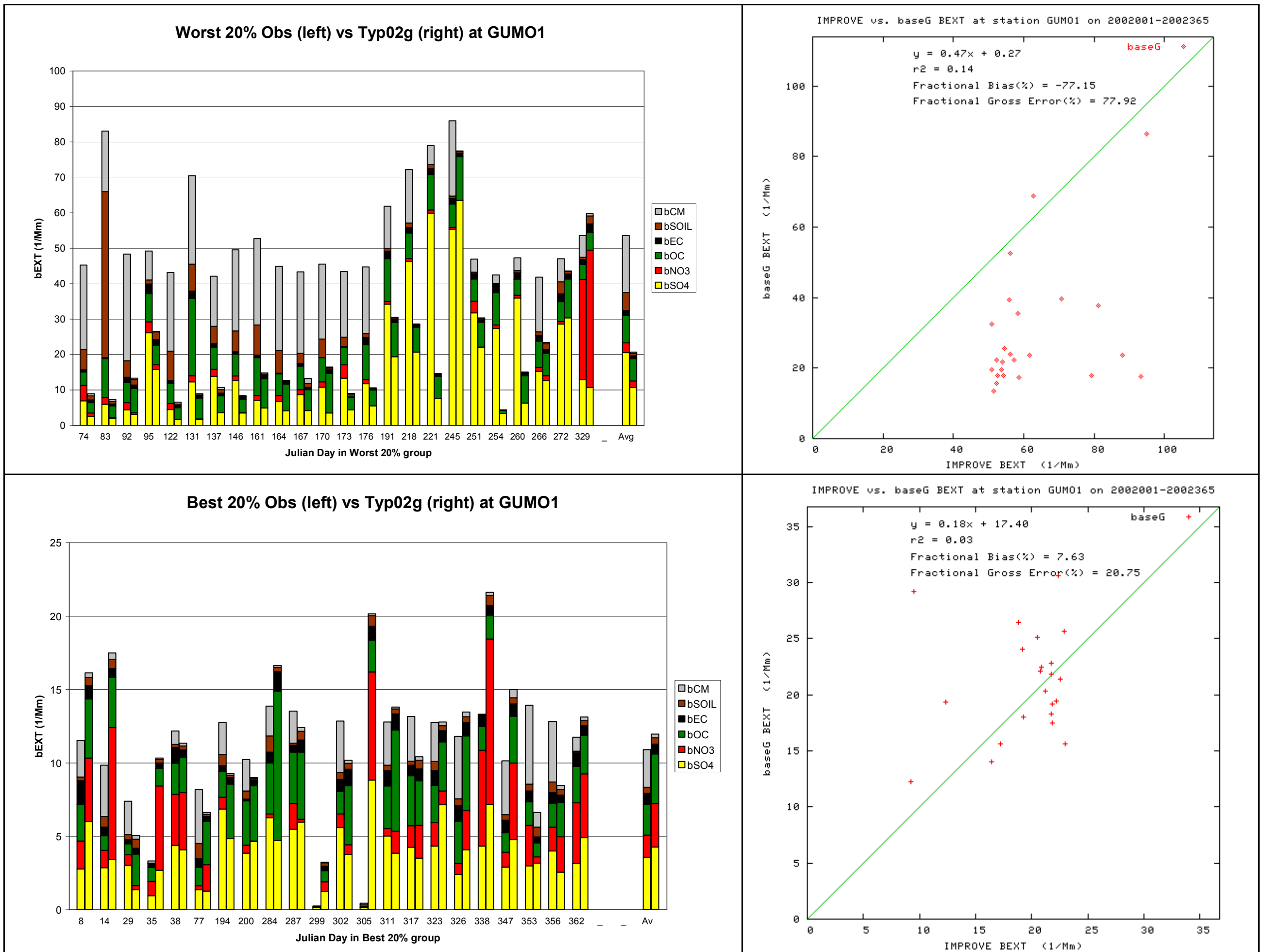


Figure 3-18. Daily extinction model performance at Guadalupe Mountains (GUMO), Texas for the worst (top) and best (bottom) 20 percent days during 2002.

3.8 Model Performance Evaluation Conclusions

The model performance evaluation reveals that the model is performing best for SO4, OMC and EC. Soil performance is mixed with winter overestimation bias but lower bias but high error in the summer. CM performance is poor year round. The operational evaluation reveals that SO4 performance usually achieves the PM model performance goal and always achieves the model performance criteria, although it does have an underestimation bias that is greatest in the summer. NO3 performance is characterized by a winter overestimation bias with an even greater summer underestimation bias. However, the summer underestimation bias occurs when NO3 is very low and it is not an important component of the observed or predicted PM and visibility impairment. Performance for OMC meets the model performance goal year round at the IMPROVE sites, but is characterized by an underestimation bias at the more urban STN sites. EC exhibits very low bias at the STN sites and a summer underestimation bias at the IMPROVE sites, but meets the model performance goal throughout the year. Soil has a winter overestimation bias that exceeds the model performance goal and criteria raising questions whether the model should be used for this species. Finally, CM performance is extremely poor with an under-prediction bias that exceeds the performance goal and criteria. We suspect that much of the CM concentrations measured at the IMPROVE sites is due to highly localized emissions that can not be simulated with 36 km regional modeling.

Performance for the worst 20 percent days at the CENRAP Class I areas is generally characterized by an underestimation bias. Performance at the BRET, BIBE and GUMO Class I areas for the worst 20 percent days is particularly suspect and care should be taken in the interpretation of the visibility projections at these three Class I areas.

The CMAQ 2002 36 km model appears to be working well enough to reliably make future-year projections for changes in SO4, NO3, EC and OMC at the rural Class I areas. Performance for Soil and especially CM is suspect enough that care should be taken in interpreting these modeling results. The model evaluation focused on the model’s ability to predict the components of light extinction mainly at the Class I areas. Additional analysis would have to be undertaken to examine the model’s ability to treat ozone and fine particulate to address 8-hour ozone and $PM_{2.5}$ attainment issues.